Self-Refining Vision Language Model for Robotic Failure Detection and Reasoning

Carl Qi[1], Xiaojie Wang[2], Silong Yong[3], Stephen Sheng[2], Huitan Mao[2], Sriram Srinivasan[2], Manikantan Nambi[2], Amy Zhang[1], Yesh Dattatreya[2]
[1]UT Austin, [2]Amazon Robotics, [3]Carnegie Mellon University

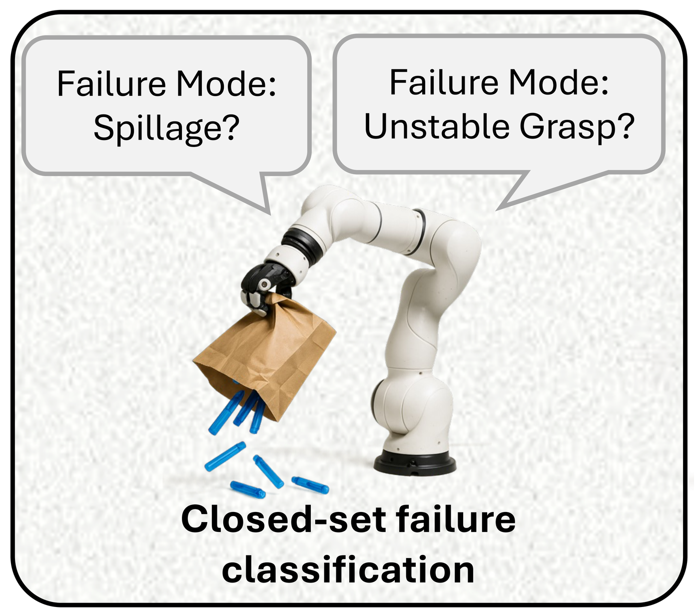
The University of Texas at Austin
Cockrell School of Engineering

amazon robotics

Carnegie Mellon University

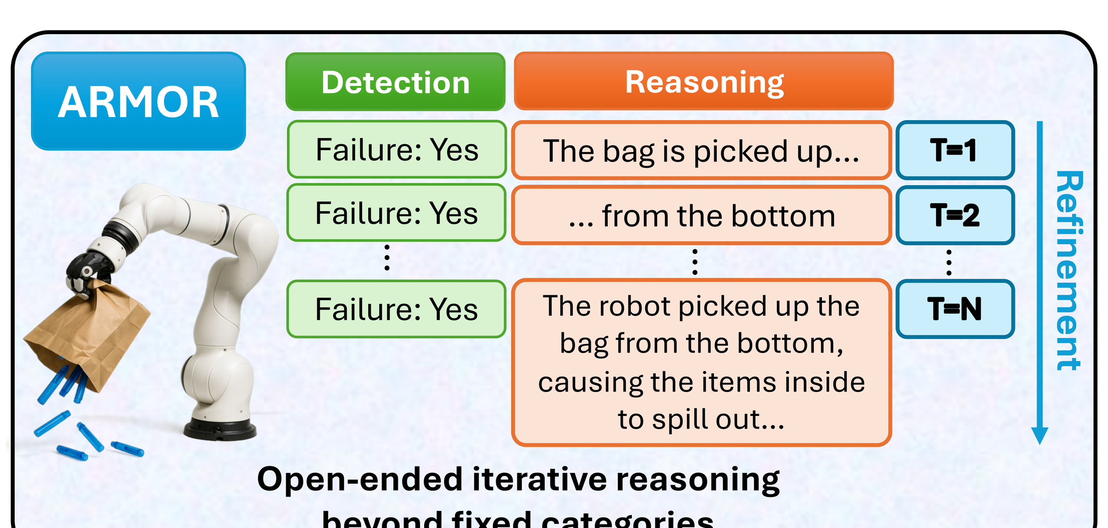
Introduction

Robotic systems fail in the real world in ways that are **subtle**, **combinatorial**, and **difficult to enumerate**. Prior approaches treat failure reasoning as a closed-set classification problem that cannot generalize beyond predefined failure modes. However, rich human annotations are expensive to acquire. There is a clear need for a system that can perform open-ended failure detection and reasoning under limited supervision.

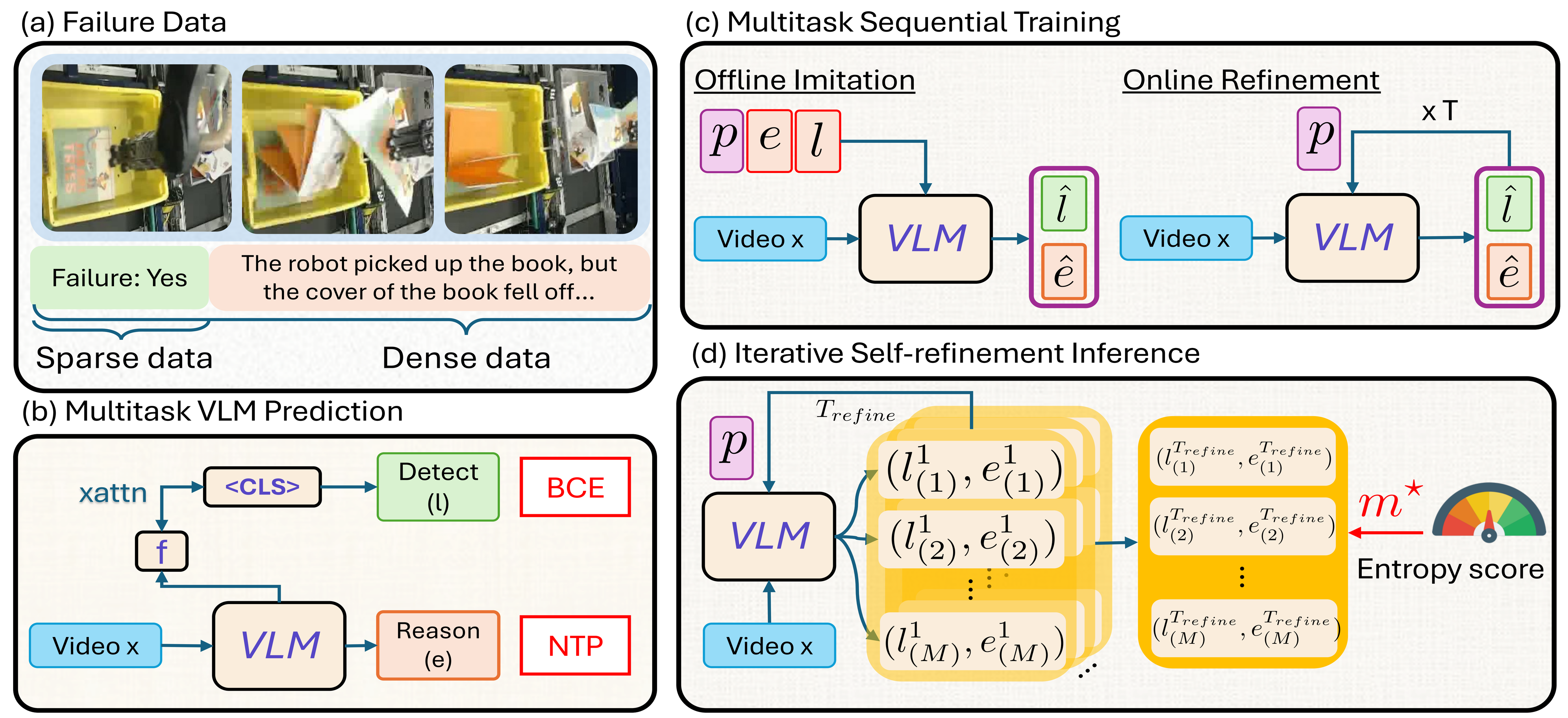


Research Goal

We aim to build a unified model that jointly detects robotic failures and generates rich natural language explanations beyond predefined failure categories.
We focus on a heterogeneous supervision: large-scale sparse binary detection labels and small-scale rich reasoning annotations.



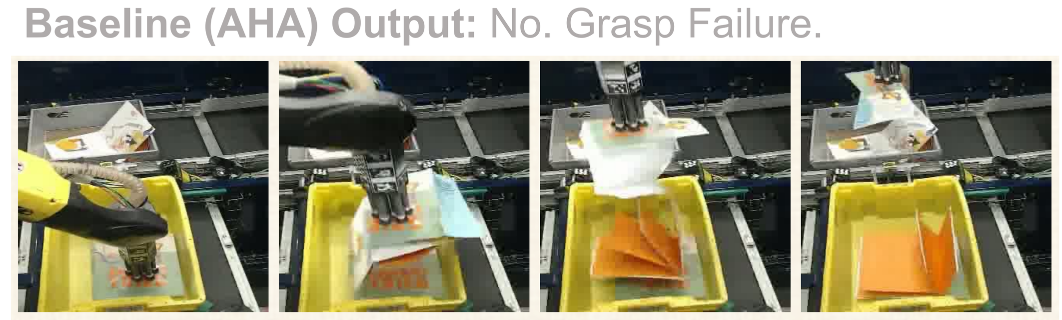
Methods



We introduce **ARMOR** (Adaptive Round-based Multi-task mOdel for Robotic failure detection and reasoning), a self-refining VLM. (a) We consider large amount of binary detection labels and scarce free-form reasoning labels. (b) We adapt a VLM to have multiple prediction heads using the shared representation. (c) We finetune the VLM with both offline BC and to refine its own outputs in an online fashion. (d) At inference, we roll out multiple refinement turns and pick the final answer associated with the lowest entropy.

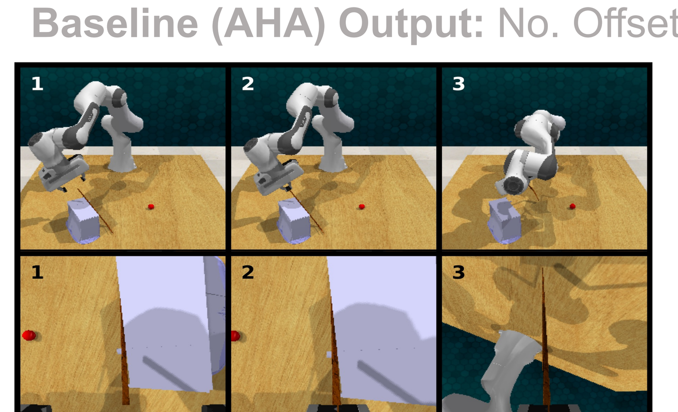
Figures and Results

Multistep reasoning qualitative example 1:
Q: In the video, the robot is attempting to transport an object from one bin to another. Did the robot successfully transport the object?
Baseline (AHA) Output: No. Grasp Failure.



ARMOR Refinements (R: refinement round):
R1: I observe the robot's arm or gripper making an unintended forceful contact with objects in the environment. The impact is clearly visible as the robot's smooth trajectory is disrupted when it strikes either the tote wall, other items, or surrounding structures
R3: I see that the robot has successfully picked up the book from the source location, but the cover of the book fell off, causing damage to the item and failure to the operation.

Multistep reasoning qualitative example 2:
Q: In the video, the robot is attempting to pick up a hockey stick. Did the robot succeed at the subtask?
Baseline (AHA) Output: No. Offset along the y-axis.

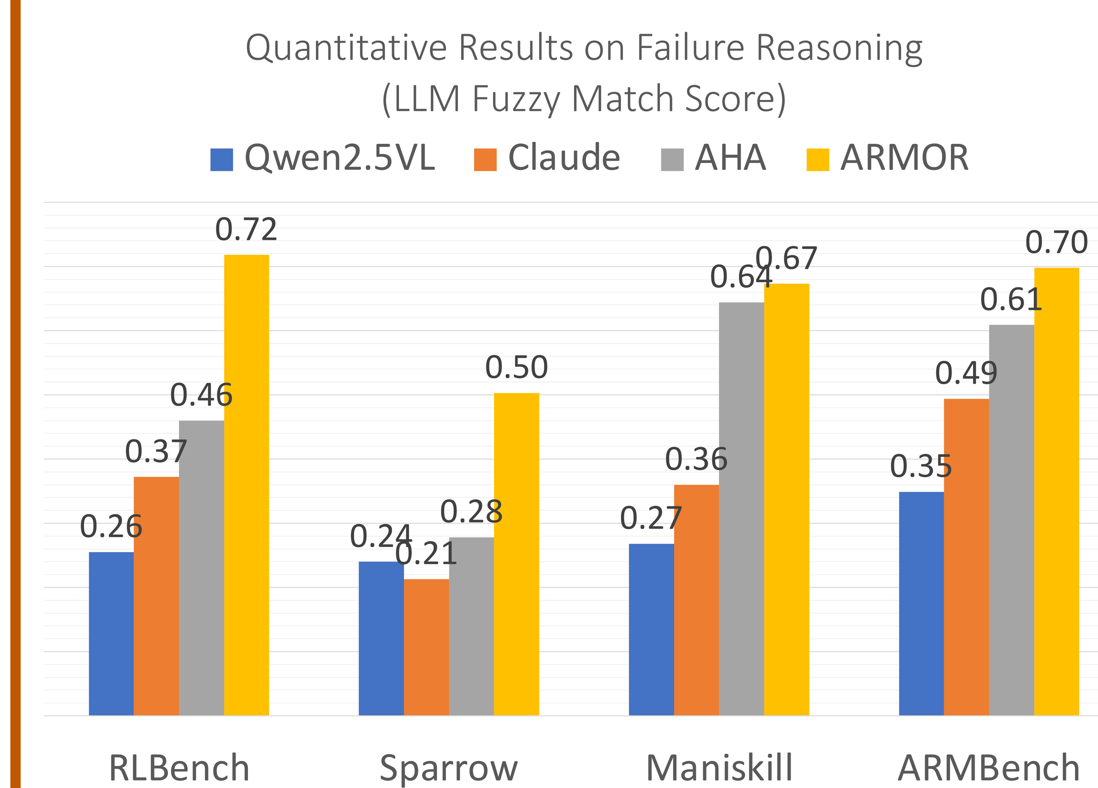


ARMOR Outputs (R=round outputs):
R1: No, The robot gripper moved with an offset along the y direction.
R2: Yes, The robot fails to close the gripper.
R3: Yes, The robot succeeded at the sub-task.

Ablation	Offline Warmup	Offline Expert Condition	Online Imitation	Refinement (Inference)	Detection / Reasoning
Multitask Prediction	✓	✗	✗	✗	0.897 / 0.460
Refinement Only	✓	✗	✗	✓	0.803 / 0.488
Offline Imitation Only	✓	✓	✗	✓	0.853 / 0.658
Online Imitation Only	✗	✗	✓	✓	0.850 / 0.683
ARMOR (ours)	✓	✓	✓	✓	**0.917 / 0.718**

Ablation study results showing the contribution of each component in ARMOR.

Conclusion



ARMOR demonstrates that robotic failure detection and reasoning can be unified in a self-refining multi-task framework. ARMOR achieves SoTA performance with up to **30%** improvement in detection and **100%** improvement in reasoning. Future work includes **incorporating modalities beyond vision**, including force torque sensing and proprioception as well as extending to **failure recovery**.

Acknowledgments

The project was developed during CQ's internship at Amazon Robotics. AZ and CQ were funded by NSF 2340651, NSF 2402650, DARPA HR00112490431, and ARO W911NF-24-1-0193.

References

Duan et al. AHA: A Vision-Language-Model for Detecting and Reasoning over Failures in Robotic Manipulation. ICLR 2025.
Madaan et al. Self-Refine: Iterative Refinement with Self-Feedback. NeurIPS 2023.
Ross et al. A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning. AISTATS 2011.
Kang et al. Scalable Best-of-N Selection for Large Language Models via Self-Certainty. arXiv 2025.